

ON-LINE DICTIONARY AND READ UNDERSTANDING SUPPORT SYSTEM UTILIZING SAME

BEST AVAILABLE COPY

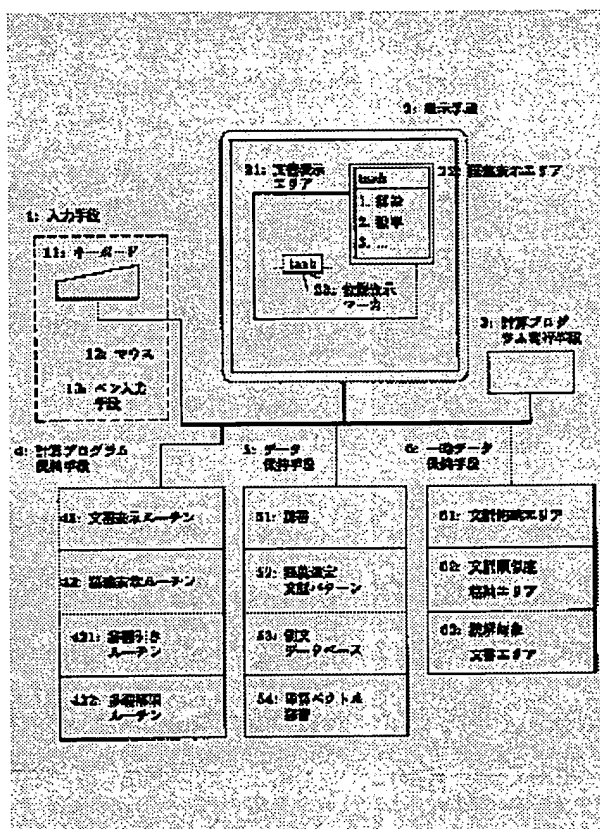
Patent number: JP8235181
Publication date: 1996-09-13
Inventor: NIWA YOSHIKI; NITTA YOSHIHIKO
Applicant: HITACHI LTD
Classification:
 - international: G06F17/27; G06F17/28
 - european:
Application number: JP19950039673 19950228
Priority number(s):

Report a data error here

Abstract of JP8235181

PURPOSE: To speedily know a proper equivalent with small labor by displaying a proper equivalent of the context to a specified word in easy-to-see form for a user at the top or upper part of a synonym display part or by coloring, etc.

CONSTITUTION: When the user specifies a document to be read through an input means 1, a computation program execution means 3 displays the necessary document in a document display area 21 in a display means 2 by using a document display routine 41 and the user beings to read it. If a word whose meaning is unknown is found, the word on the screen is specified with a pen input means 13. The computation program execution means 3 actuates a synonym display routine 42 in response to the specification of the word to open a synonym display area 22 on the screen nearby the word, and the meaning of the word stored in a dictionary 51 is displayed on the screen. At this time, when the word has plural meanings, the meaning matching the context of the passage is preferentially displayed at an upper part in consideration of the context.



Data supplied from the esp@cenet database - Patent Abstracts of Japan

THE PAGE BLANK (USPTO)

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平8-235181

(43)公開日 平成8年(1996)9月13日

(51)Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 17/27		8420-5L	G 0 6 F 15/38	M
17/28		8420-5L		T

審査請求 未請求 請求項の数 2 O L (全 7 頁)

(21)出願番号 特願平7-39673

(22)出願日 平成7年(1995)2月28日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 丹羽 芳樹

埼玉県比企郡鳩山町赤沼2520番地 株式会

社日立製作所基礎研究所内

(72)発明者 新田 義彦

埼玉県比企郡鳩山町赤沼2520番地 株式会

社日立製作所基礎研究所内

(74)代理人 弁理士 小川 勝男

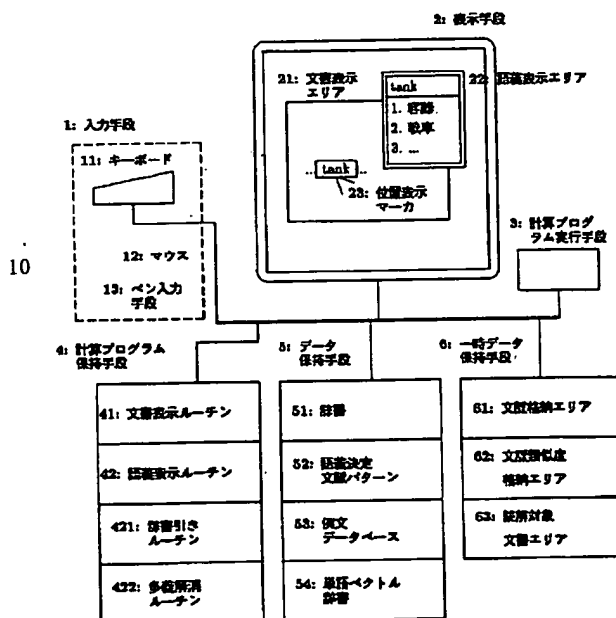
(54)【発明の名称】 オンライン辞書およびこれを利用した読解支援装置

(57)【要約】

【目的】 文脈に応じた適訳を語義表示部の先頭に表示することにより少ない労力でかつ迅速に適訳を知ることができるようにすること。

【構成】 入力手段1、表示手段2、計算手段3、および記憶手段4、5、6を有する計算機上で稼働する。辞書引きする単語に多義性がある場合には多義解消ルーチン422が語義決定文脈パターン52と例文データベース53および文脈間の類似度を判定するルーチンなどを用いて当該文脈に置ける単語の意味を判定する。

図 1



1

【特許請求の範囲】

【請求項 1】文脈中の単語の意味を判定するために、語義判定用文脈パターンと文脈例のデータベース、および文脈間の類似度を判定するルーチンを持ち、始めに語義判定用文脈リストと当該文脈の照合を試み、適合するものがあればそれに対応する語義であると判定し、それが無い場合には文脈例と当該文脈との類似度を概類似度判定ルーチンにより判定し最も類似度の高い文脈に置ける意味であると判定する多義解消機構を有するとともに、適訳に近いものを優先的に又は他の表示と異なる表示によ

って表示するオンライン辞書。
【請求項 2】読解すべき文章を記憶する手段を備え、前記文章中の特定の単語に関し、文脈中の単語の意味を判定するために、語義判定用文脈パターンと文脈例のデータベース、および文脈間の類似度を判定するルーチンを持ち、始めに語義判定用文脈リストと当該文脈の照合を試み、適合するものがあればそれに対応する語義であると判定し、それが無い場合には文脈例と当該文脈との類似度を概類似度判定ルーチンにより判定し最も類似度の高い文脈に置ける意味であると判定する多義解消機構を有するオンライン辞書を使用した読解支援装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明はオンラインの文書をディスプレイ上で読む時に利用する辞書およびこれを利用した読解支援装置に関する。

【0002】

【従来の技術】従来の電子辞書は指定した単語に関する記載事項を固定的に表示する。

【0003】

【発明が解決しようとする課題】単語は一般に複数の意味を持ち、それらは文脈に依存する。単語に関する固定的な辞書記載事項を表示するのでは、当該文脈における最適の意味が先頭(もしくは上位)に表示されるとは限らないので、場合によっては適訳を探し当てるまでに時間と労力がかかることがある。

【0004】

【課題を解決するための手段】文脈中の単語の意味を判定するために、語義判定用文脈パターンと文脈例のデータベース、および文脈間の類似度を判定するルーチンを持ち、始めに語義判定用文脈リストと当該文脈の照合を試み、適合するものがあればそれに対応する語義であると判定し、それが無い場合には文脈例と当該文脈との類似度を概類似度判定ルーチンにより判定し最も類似度の高い文脈に置ける意味であると判定する多義解消機構を備える。

【0005】

【作用】本発明では、文脈に応じて単語の意味の優先度を決定する手段を利用することにより、指定された単語の当該文脈における最適の意味が先頭(もしくは上位)に

2

表示される。したがって、指定した単語の当該文脈における適訳が語義表示部の先頭もしくは上位に表示されるので、少ない労力、例えば語義表示部をスクロールする量が少ない等、かつ迅速に適訳を知ることができる。さらには、読書の効率化を計ることが出来る。

【0006】

【実施例】以下、本発明による文脈考慮型のオンライン辞書の構成および使用の一実施例を図1に従って説明する。

【0007】本実施例では入力手段1、表示手段2、計算プログラム実行手段(マイクロプロセッサ)3、計算プログラム保持手段4、データ保持手段5、一時データ保持手段(半導体メモリなど)6を有する計算機システム上で稼働する。入力手段1は例えば、キーボード11、マウス12あるいはペン入力手段13よりなる。表示手段2は、後述するように各種の計算プログラムの機能が發揮されるのに対応して、必要な表示エリア、例えば文書表示エリア21、語義表示エリア22を形成され、位置表示マーカ23が表示される。、計算プログラム実行手段(マイクロプロセッサ)3は計算プログラム保持手段4に保持される各種ルーチン、例えば、文書表示ルーチン41、語義表示ルーチン42、辞書引きルーチン421、多義解消ルーチン422を所定のプログラムに従って実行する。データ保持手段5には、例えば、見出し語と意味を対として保持する通常の電子辞書51、語義決定文脈パターン52、例文データベース53、単語ベクトル辞書54が保持される。一時データ保持手段6には、例えば、文脈格納エリア61、文脈類似度格納エリア62、読解対象文書エリア63が形成される。

【0008】利用者は読解の対象となる文書の内容を、図の入力手段1あるいは電子データとして図示されていない入力手段を使用して読解対象文書エリア63に保持する。利用者が読解の対象となる文書を入力手段1で指定すると、計算プログラム実行手段3は文書表示ルーチン41を用いて表示手段2の中の文書表示エリア21に必要な文書が表示され利用者はこれを読み始める。その際に意味が良く分からない単語に遭遇した場合、画面上の該当単語をペン入力手段13で指定したり、またはマウス12などの入力手段を用いて位置表示マーカ23を該当する単語上に持っていく、クリックなどの操作により指定する。計算プログラム実行手段3は単語の指定に対応して、語義表示ルーチン42を起動し語義表示エリア22を当該単語の近辺の画面上に開き、辞書51に記憶されている単語の意味がその画面上に表示される。この際、本発明では、その単語が複数の意味を持つ場合には、前後の文脈を考慮に入れることによりその文脈にふさわしい意味を優先的に上位に表示することにより、利用者が適訳を選択する労力を軽減する。その単語の意味が一つの場合にはそれを表示するのは当然である。

【0009】以下、この文脈を考慮に入れて適訳を優先

50

3

的に表示する作業をいくつかの例を用いて説明する。

【0010】図3は「...with fuel tank capable of...」という文脈で使われている tank の辞書引きを行なった場合である。該当単語に位置表示マーカ23を当て単語を指示すると、始めに辞書引きルーチン421により tank の語義が引かれる。

【0011】図2に tank の語義を示す。この辞書の例では、'tank' に関する記載511は、意味1から意味3までの語義を持つものとして定義されている。これから tank が複数の意味を持つことが分かり、多義解消ルーチン422が起動される。図4はこの多義解消ルーチン422の構成を示したものであり、図に示すように、文脈切り出しルーチン4221、語義決定文脈パターンとの照合ルーチン（以下単に照合ルーチンと呼ぶ）4222、および例文との類似度計算ルーチン4223とよりなる。多義解消ルーチン422は、始めに文脈切り出しルーチン4221を起動して、指定された単語 tank の前後の文脈を一時データ保持手段6の中の文脈格納エリア61に格納する。図5はその実行後の同エリアの例である。左側のコラムの数字は指定された単語との相対的な位置関係を示し、例えば「-2」の欄には指定された単語の2つ手前の単語「with」が格納される。

【0012】続いて多義解消ルーチン422は照合ルーチン4222を起動する。照合ルーチン4222は語義決定文脈パターン52の中の tank に関するパターン（図6）と、先に文脈格納エリア61に格納された、図5に示す tank に関する文脈との照合を行なう。

【0013】ここで、語義決定文脈パターン52を図6を例に説明する。

【0014】語義決定文脈パターンは定型句（イディオム）として知られているものの他、新聞などの大量の文書データから頻出するパターンを抽出して作成する。すなわち、図6の例では think tank と tank top は通常イディオムとして知られているものであり、fuel tank 途ant i tank missail は新聞などの文書データに頻出するものを収集したものである。

【0015】この場合では、図5に示す tank に関する文脈は、語義決定文脈パターン52の上から3番目のパターン「fuel tank」に適合するので、その場合の意味番号'tank'に関する辞書51の記載511の1番の意味が当該文脈における当該単語の意味であると判定する。

【0016】この例では、多義解消ルーチン422中の例文との類似度計算ルーチン4223は使わなかったが、これは語義決定文脈パターンの中に適合するものが無い場合に使用される。その例は利用例の3として後に示す。

【0017】上記の判定に従って、図7に示すように、語義表示ルーチン42は tank の意味のうち1番の意

4

味「容器」を語義表示エリア22の先頭に表示する。それ以外の意味も辞書記載の順に表示する。なお第1候補の意味を決定するのに寄与した文脈、この場合では「fuel tank」、は利用者にとって有益な情報なので、それを明示するためアンダーラインなどの文脈表示マーカ24を表示する。最適な意味として判定された意味の語義表示エリアにおける表示は、上記のように、最上段に表示するほか、順序は変えず、着色もしくは字種変更等、他の表記と区別しやすい異なった表記で強調してして表示するのも良い。

【0018】図8(a)は「...Mexico City think tank, says ...」という文脈で使われている tank の辞書引きを行なった場合である。前の例と同様の動作をし、文脈格納エリア61に図8(b)のような文脈が格納される。続いて照合ルーチン4222は語義決定文脈パターン52中の tank に関するパターン（図9

(b)）と当該文脈（図8(b)）の照合を行なう。この場合では一番上のパターンに適合するが、この場合には意味記述欄には「Idiom」と記されており、これはイディオムとして辞書中に tank とは別の見出しとして登録されていることを示す。従って辞書引きルーチン421が起動されて、図5に示す辞書51中の「think tank」に関する記載事項が図9上(a)のように語義表示エリア22に表示される。この場合にも意味を決定するのに寄与した文脈（イディオム）、「think tank」、に文脈表示マーカ24が表示される。

【0019】図10(a)は「...for tank ammunition and ...」という文脈で使われている tank の辞書引きを行なった場合である。前の例と同様の動作をし、文脈格納エリア61に図10(b)のような文脈が格納される。続いて照合ルーチン4222は語義決定文脈パターン52中の tank に関するパターン（図9(b)）と当該文脈（図10(b)）を上位の規則から順に照合する。この場合には適合するパターンが存在しないので、このままでは意味の判定ができない。そこで多義解消ルーチン422は、さらに、図4に示す「例文との類似度計算ルーチン」4223を起動し、当該文脈と最も類似した例文から意味を推定する。

【0020】例文は「例文データベース」53に蓄えられている。図11は例文データベース中の「tank」に関する部分531の一例を示したものである。各例文データはどの単語に関する例文であるかを示す部分（一番左）と例文の番号、その例文における当該単語の意味の番号（利用する辞書51に記述されている意味番号）からなっている。

【0021】類似度計算ルーチン4223はこれらの例文と当該文脈（図10(b)）の類似度を計算しその結果を文脈類似度格納エリア62に順次格納する。図12はその結果の一例である。この結果から最も類似度の高

5

い例文を選択する。この場合には5番の例文が一番類似度が高いので、例文データベースの tank に関する例文5 (図11の5311) の意味番号2がこの場合の tank の意味として判定される。従って、図2に示す辞書中の tank に関する記述511の意味番号2に相当する部分「戦車」が、図13に示すように、語義表示エリア22の先頭に表示される。その他の意味は辞書見出しの順で記述される。この場合には意味を決定するのに寄与した文脈の範囲は特定できないので文脈表示マークは表示されない。

【0022】文脈間の類似度を測る方法は色々であるが、ここでは文脈を文脈ベクトルというベクトル(数値列)の形に変換し、それら数値列間の類似度(ベクトルのなす角度)で測る方法を示す。図14に示したように「例文との類似度計算ルーチン」4223は「文脈ベクトル作成ルーチン」42231と、「文脈ベクトルの類似度計算ルーチン」42232とからなる。

【0023】文脈ベクトルの作成方法は色々であるが、ここではその一例として文脈を構成する単語のベクトルの和を取る方法を用いる。図15に示したように、単語ベクトル辞書54には各単語の意味を反映すると考えられる一定の長さの数値列(単語ベクトル)が格納されている。単語ベクトルの作り方には色々あるが、一例としては特開平6-332935号「単語の数値列表現辞書、その作成方法およびそれを使用した装置」に示されている方法で作成することができる。

【0024】一般に単語wの文脈C(後方にn単語までと前方にm単語まで)を「w(-n)・・・w(-2) w(-1) w w(1) w(2)・・・w(m)」とする。ここでw(-i)は単語wから見て後方へi番目の単語w(i)は前方へi番目の単語を示す。このとき、この文脈Cの文脈ベクトルはこれらの単語のベクトルの和で計算される。ただしw自身の単語ベクトルは加えないものとする。すなわち、単語wの単語ベクトルをWV[w]と書くことにすると、上記の文脈のベクトルCV[C]は

【0025】

【数1】

(数1)

$$CV[C] = \sum_{i=-n}^{-1} WV[w_i] + \sum_{i=1}^m WV[w_i]$$

【0026】で計算される。

【0027】このような計算を実行する文脈ベクトル作成ルーチン42231を用いて現在着目している文脈(図10(b)、C0と表示する)と各文脈例(図9(b)、CE1, ..., CE5と表示する)の文脈ベクトルを作成する。すなわちCV[C0]とCV[CE1], ..., CV[CE5]が計算される。そこで当該文脈のベクトルCV[C0]と各文脈例の文脈ベクトル

6

の間の類似度を「文脈ベクトルの類似度計算ルーチン」42232が計算して、文脈の類似度格納エリア62に格納する。一般に二つの文脈ベクトルの類似度はそれらのなす角度のコサインで計算される。従って二つの文脈ベクトル(U, V)の間の類似度は

【0028】

【数2】

(数2)

$$\text{類似度}(U, V) = \frac{\sum_{i=1}^N u_i * v_i}{\left(\sum_{i=1}^N u_i^2\right)^{1/2} * \left(\sum_{i=1}^N v_i^2\right)^{1/2}}$$

ここで $U = (u_1, \dots, u_N),$

$V = (v_1, \dots, v_N).$

【0029】で計算される。

【0030】

【発明の効果】本発明により、指定した単語の当該文脈における適訳が、語義表示部の先頭もしくは上位、あるいは着色等の、使用者に見やすい形に表示されるので、少ない労力で(例えば語義表示部をスクロールする量が少なく済むように)かつ迅速に適訳を知ることができるようになる。

【図面の簡単な説明】

【図1】電子辞書の構成の一例を示す図。

【図2】辞書記載内容の一例を示す図。

【図3】利用例1の文書表示エリアにおける表示例を示す図。

【図4】多義解消ルーチンの一例を示す図。

【図5】文脈格納エリアの一例を示す図。

【図6】語義決定文脈パターンの一例を示す図。

【図7】利用例1の文書表示エリアにおける表示例に語義表示エリアを表示した例を示す図。

【図8】(a)は利用例2の文書表示エリアにおける表示例を示す図、(b)は利用例2の文脈格納エリアの例を示す図。

【図9】(a)は利用例2の文書表示エリアにおける表示例に語義表示エリアを表示した例を示す図、(b)は利用例2の語義決定文脈パターンの例を示す図。

【図10】(a)は利用例3の文書表示エリアにおける表示例を示す図、(b)は利用例3の文脈格納エリアの例を示す図。

【図11】例文データベースの一例を示す図。

【図12】例文との類似度比較の一例を示す図。

【図13】利用例3の文書表示エリアにおける表示例に語義表示エリアを表示した例を示す図。

【図14】例文との類似度計算ルーチンの一例を示す

図。

【図15】単語ベクトル辞書の一例を示す図。

【符合の説明】

1 : 入力手段、11 : キーボード、12 : マウス、
 3 : ペン入力手段、2 : 表示手段、21 : 文書表示エ
 リア、22 : 語義表示エリア、23 : 位置表示マーカ、2
 4 : 文脈表示マーカ、3 : 計算プログラム実行手段、
 4 : 計算プログラム保持手段、41 : 文書表示ルーチ
 ン、42 : 語義表示ルーチン、421 : 辞書引きルーチ

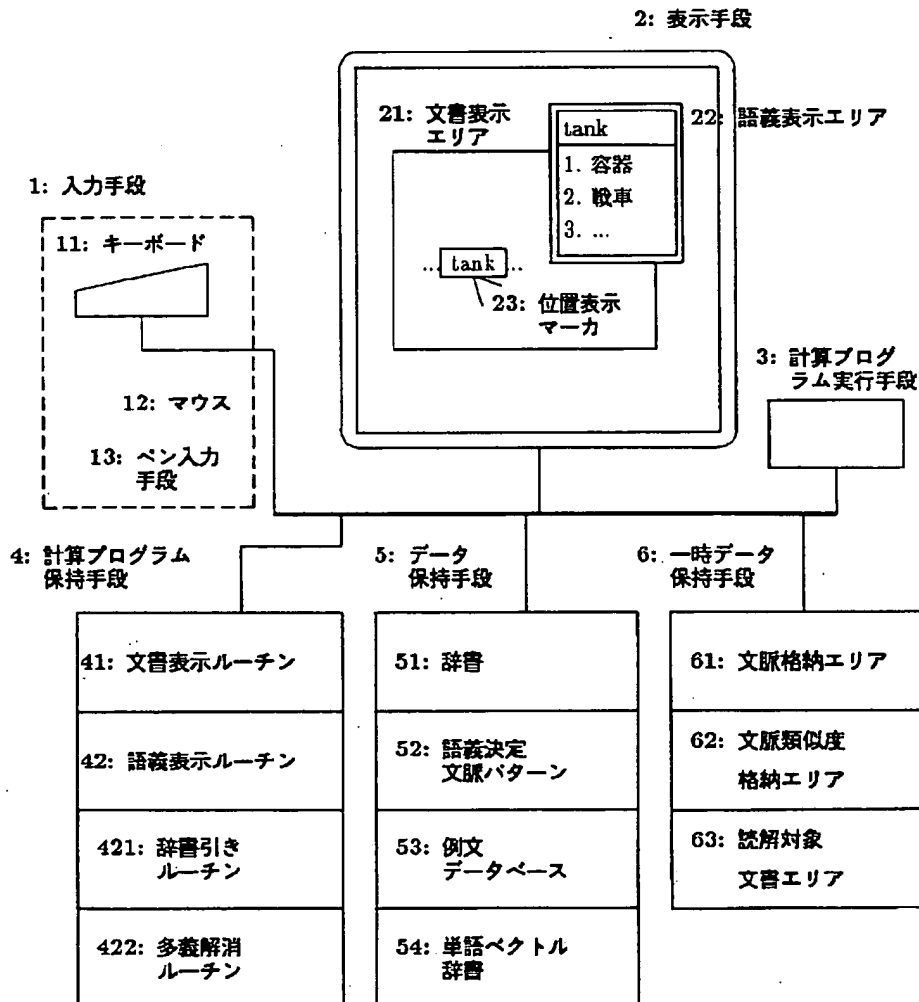
ン、422 : 多義解消ルーチン、4221 : 文脈切り出
 しルーチン、4222 : 語義決定文脈パターンとの照合
 ルーチン、4223 : 例文との類似度計算ルーチン、4
 2231 : 文脈ベクトル作成ルーチン、42232 : 文
 脈ベクトルの類似度計算ルーチン、5 : データ保持手
 段、51 : 辞書、52 : 語義決定文脈パターン、53 :
 例文データベース、54 : 単語ベクトル辞書、6 : 一時
 データ保持手段、61 : 文脈格納エリア、62 : 文脈類
 似度格納エリア。

【図1】

【図5】

図 1

図 5



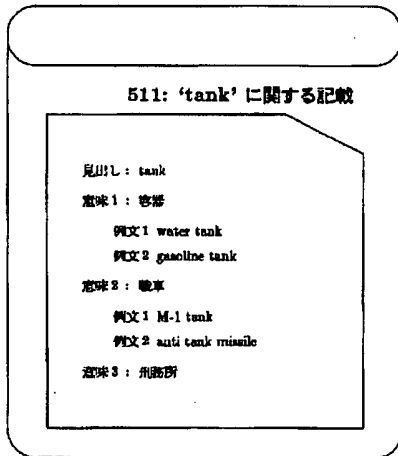
61: 文脈格納エリア

-5	the
-4	Boeing
-3	747-400
-2	with
-1	fuel
0	tank
1	capable
2	of
3	holding
4	as
5	much

【図 2】

図 2

51: 辞書



【図 6】

図 6

52: 語義決定文脈パターン

-5	-4	-3	-2	-1	0	1	2	3	4	5	意味
					think	tank					Idiom
						tank	top				Idiom
					fuel	tank					1
					anti	tank	missile				2
											:

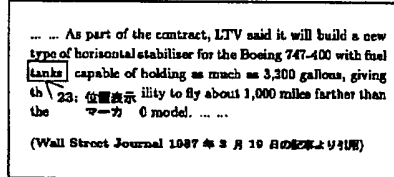
【図 8】

図 8

【図 3】

図 3

21: 文書表示エリア



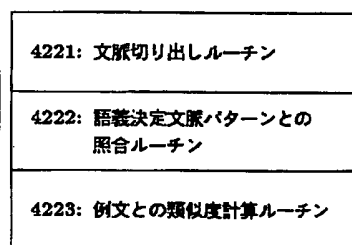
【図 7】

図 7

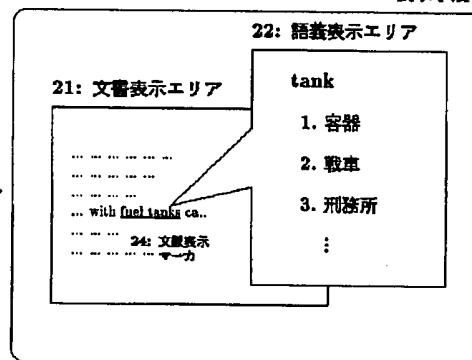
【図 4】

図 4

422: 多義解消ルーチン



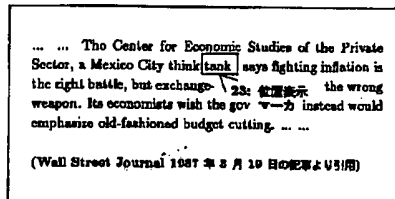
2: 表示手段



【図 10】

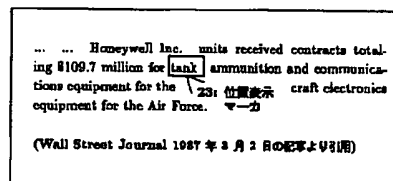
図 10

21: 文書表示エリア



(A)

21: 文書表示エリア



(A)

61: 文脈格納エリア

-5	,
-4	a
-3	Mexico
-2	City
-1	think
0	tank
1	,
2	says
3	fighting
4	inflation
5	is

(b)

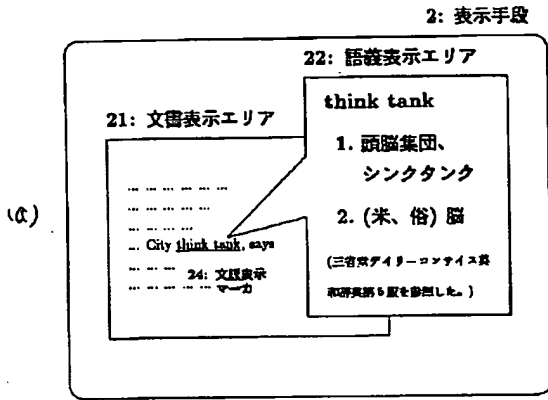
61: 文脈格納エリア

-5	contracts
-4	totaling
-3	\$109.7
-2	million
-1	for
0	tank
1	ammunition
2	and
3	communications
4	equipment
5	for

(b)

【図 9】

図 9



【図 11】

図 11

53: 例文データベース

単語	例文番号	意味番号	例文
tank	1	1	The tank will only hold three gallons
tank	2	1	There is a leak in the cold water tank
tank	3	1	a petrol tank
tank	4	2	Their equipment included more than 2,000 tanks
tank	5	2	a tank battalion

5311: tank に関する例文 5 531: tank に関する例文群

【図 14】

図 14

52: 語義決定文脈パターン

(b)

-5	-4	-3	-2	-1	0	1	2	3	4	5	意味
											:
				think	tank						Idioms
					tank	top					Idiom
				fuel	tank						1
				anti	tank	missile					2
					:						:

【図 12】

図 12

【図 13】

図 13

4223: 例文との類似度計算ルーチン

42231: 文脈ベクトル作成ルーチン

42232: 文脈ベクトルの類似度計算ルーチン

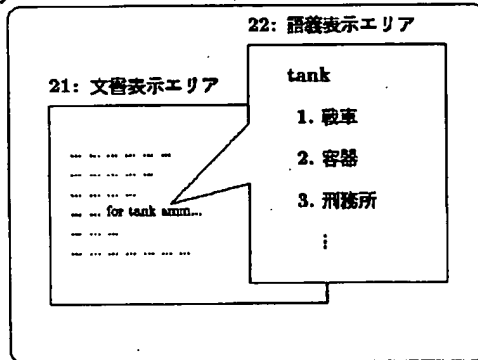
【図 15】

図 15

62: 文脈の類似度格納エリア

例文番号	類似度
1	-0.1
2	0.0
3	0.2
4	0.1
5	0.5

2: 表示手段



54: 単語ベクトル辞書

単語	第1成分	第2成分	第3成分	第N成分
:	:	:	:	:
ammunition	am ₁	am ₂	am ₃	am _N
:	:	:	:	:
and	an ₁	an ₂	an ₃	an _N
:	:	:	:	:
for	f ₁	f ₂	f ₃	f _N
:	:	:	:	:
million	m ₁	m ₂	m ₃	m _N
:	:	:	:	:

THIS PAGE BLANK (11/19/70)